

携帯端末をかざすだけで実世界とクラウドシステムをつなぐユーザーインターフェース

Attention-Focusing User Interface Using Markerless Image Recognition Technology to Connect Real World to Cloud Computing

西山 正志 大平 英貴 関根 真弘
 ■ NISHIYAMA Masashi ■ OHIRA Hidetaka ■ SEKINE Masahiro

身の回りにある実世界の物からクラウドシステム上の豊富なコンテンツへ、シームレスにつながる新しい生活スタイルに期待が高まっている。

東芝は、対象物にカメラ付き携帯端末をかざすだけで、クラウドシステム上の関連情報をすばやく取り出し提示する“かざす”ユーザーインターフェースを開発した。既に普及しているバーコードやQRコード^(注1)などのマーカでは、対象物のデザインを損ねる問題があるため、特別なマーカなしで、対象物に一致した画像が事前に登録されているか、また登録されているならそれはどれかを認識する画像認識技術を開発した。登録画像が700枚程度なら、スマートフォンやタブレットで、わずか0.5秒の間に認識率99%で対象物を見分けることができる。

Expectations are rising for the realization of a new lifestyle providing a seamless connection between real objects and a wide variety of contents accessible by cloud computing.

Toshiba has developed an attention-focusing user interface (UI) that offers a new service for connecting real objects to digital data on the Internet. The attention-focusing UI is required to quickly and correctly recognize objects around users. However, conventional technologies that achieve image recognition of an object without markers such as a bar code or QR code have several problems, such as the high computational cost and long processing time required to extract the features of the object from the image. To overcome these problems, we have developed a new markerless image recognition technology for extracting region-based edge features that can quickly and robustly recognize objects and extract the registered data. We have conducted verification tests using smartphones and tablets and confirmed that an object recognition accuracy of 99% is achieved with a response time of 0.5 seconds.

1 まえがき

近年、身の回りで気になった実世界の物にデジタル機器をかざして、関連する情報をその場で手軽に入手する新たな生活スタイルに期待が高まっている。例えば、雑誌にスマートフォンやタブレットをかざすことで、事前に登録しておいた動画を再生したり、Webサイトにつないだりすることによって、ユーザーにこれまでにない驚きや、おもしろさ、手軽さなどを与えることができると考えられる。リテール業の分野では、O2O (Online to Offline) とも言われ、インターネットでの購買活動と実店舗での購買活動を連動させ相乗効果を高めることが期待されている。将来的には、例えば図1のように、友達が着ている服が気になったら、その服にスマートフォンをかざすことで、関連するお勧め商品情報をクラウドシステムから即座に提示するという、これまでにないショッピングスタイルが考えられる。このようなO2Oに向けたサービスを実現するためには、携帯端末に搭載されたカメラで撮影した画像から、対象物を正しく認識する画像認識技術が必須になる。

カメラ画像を用いた認識技術として、従来はバーコードやQR

(注1) 白と黒の格子状のパターンで情報を表すマトリックス型2次元コードの一種で、(株)デンソーウェーブの登録商標。



図1. かざすユーザーインターフェースの狙い — ユーザーが身の回りの対象物に携帯端末をかざすと、事前に登録した関連データが提示されることを目指している。

Target of attention-focusing UI

コードなどのマーカを利用する手法が一般的であった。しかし、無機質なマーカは見た目が悪くなるなどデザイン性を損ねるとい問題があるため、マーカレスの画像認識技術が必要になる。しかし、これまでの画像認識技術は、計算量が多く、処理コストが高いという課題があり、大量の対象物を認識しようとする携帯端末では動作が遅くなって快適ではなくなるため、ユーザーにとっては不便なインターフェースとなってしまう。

そこで東芝は、認識精度が高く、かつ処理コストも低い画像認識技術の確立を目指して、対象物に携帯端末をかざすだけで、クラウドシステム上の関連情報をすばやく提示する、かざすユーザーインタフェースを開発した。

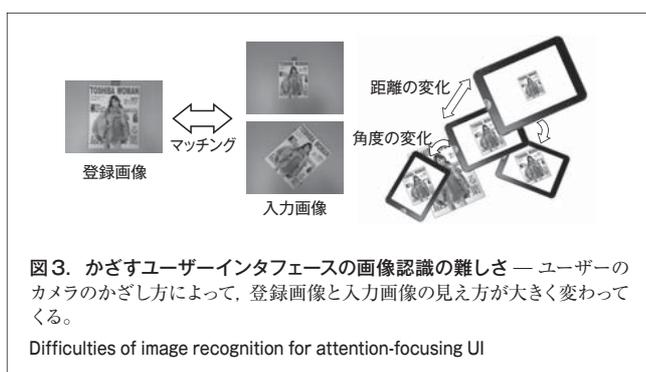
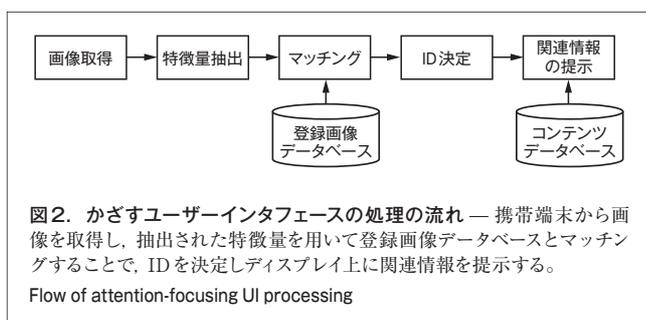
ここでは、かざすユーザーインタフェースのコア技術であるカメラを用いた画像認識技術について述べる。

2 かざすユーザーインタフェースの画像認識技術

2.1 概要

画像認識技術を用いたかざすユーザーインタフェースの処理の流れを図2に示す。携帯端末のカメラで取得した入力画像から対象物の認識に用いる特徴量を抽出し、この特徴量と登録画像データベースとをマッチングすることで、入力画像中に存在する対象物のID (Identification) を決定する。このIDに関連付けられたコンテンツを、ユーザーの携帯端末のディスプレイ上に関連情報として提示する。

このインタフェースをユーザーにとって使い心地が良いものにするためには、応答性が良く、かつ対象物を正しく認識することが重要になる。しかし、登録画像と入力画像に映っている対象物の見え方は、ユーザーのカメラのかざし方によって大きく変わってくるため、対象物を正しく認識することは難しくなる。例えば、図3のように対象物とカメラの角度や距離が変わると、画像中の対象物の見え方は大きく変わってしまう。更に、照明変動の影響も問題となってくる。すばやく、かつ画像から対象物を正しく認識するためには、認識に用いる特徴量



がもっとも重要なポイントとなる。撮影した画像中で対象物の見え方が変わっても、同様の特徴量を抽出することができれば、認識精度は高くなる。

画像認識の分野では、従来から用いられている特徴量としてエッジを用いた手法^{(1), (2)}が広く知られている。エッジは画像中の輪郭やコーナーに対応し、画像の画素値をそのまま利用した場合に比べて、照明変動に対してロバストに対象物を認識できる利点がある。一般の物体認識に向けて開発された特徴量であるORB (Oriented FAST (Features from Accelerated Segment Test) and Rotated BRIEF (Binary Robust Independent Elementary Features)) は、角度や距離の変化の影響を受けにくい画像中の局所的な領域のエッジを用いることで、高い認識率を実現している。しかし、多数の局所領域どうしをマッチングする必要があるため、登録画像の枚数が増加するとマッチングの応答速度が低下するという課題がある。また、人物検出で広く用いられている特徴量であるHOG (Histograms of Oriented Gradients) は、人などの対象物の領域を局所ではなく一つの領域として特徴量を定義している。そのため、登録画像の枚数が増加しても、マッチングに時間が掛からないという利点がある一方で、角度や距離の変化に認識精度が影響を受けやすいという欠点がある。

これらの課題に対して、当社は、良好な応答速度と高い認識精度を実現できる領域ベースエッジ特徴量を開発した。

2.2 領域ベースエッジ特徴量

エッジは隣接画素の輝度差を使って算出することが一般的である。しかし、この隣接画素を用いる方法は、中心点とその隣り合う点だけの画素情報しかエッジに含まれないため、登録時と認識時で画像上の対象物の位置が少しでも移動するとエッジの値が大きく変化してしまう。したがって、その対象物を正しく表す特徴量にならず、対象物を認識できなくなるという課題がある。その例を図4(b)に示す。隣接画素との輝度差を用いる従来のHOGは、エッジを粗く量子化することで位置ずれを許容することを狙っているが、量子化を粗くしすぎると位置ずれの問題は軽減するが、他の対象物と区別できなくなるという課題が新たに発生する。HOGはこれらの課題に対応するために画像から多数の点をサンプリングし、そのサンプリング点の隣接画素からエッジを算出しているが、サンプリングに時間が掛かり特徴量を抽出する応答速度が低下するという課題がある。

これらの課題に対し、従来の点と点との関係ではなく、面と面との関係からエッジを算出することで、位置ずれが生じた場合でもロバストに対象物を認識できる領域ベースエッジ特徴量を開発した。この手法は、離れた位置に配置した複数の周辺領域から、それぞれの周辺領域内の輝度値合計を算出した後、周辺領域どうしの輝度値合計の差を取ることでエッジを算出する。その例を図4(a)に示す。周辺領域を用いることで、

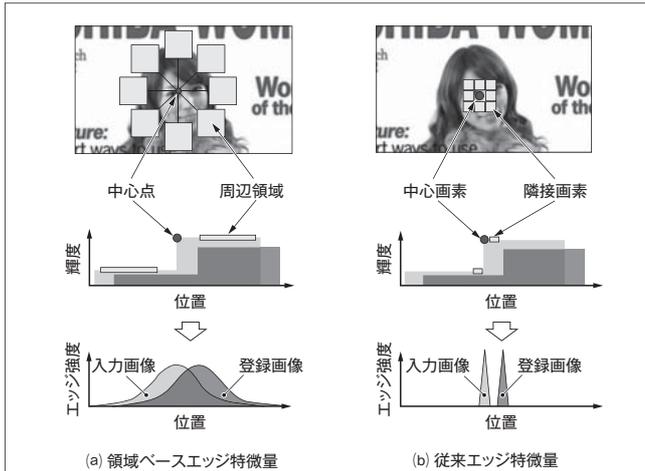


図4. 領域ベースエッジ特徴量と従来エッジ特徴量の違い — 開発した領域ベースエッジ特徴量は、従来の隣接画素から求めるエッジ特徴量に比べてより広い範囲の画像情報を用いる。
Difference between conventional edge and region-based edge feature quantities

エッジに広い範囲の画素情報を持たせることが可能になり、位置ずれに対してロバストになると考えられる。更に、この領域ベースエッジ特徴量を用いることで、図5の予備実験の結果が示すように、エッジのサンプリング回数が大幅に減少し、

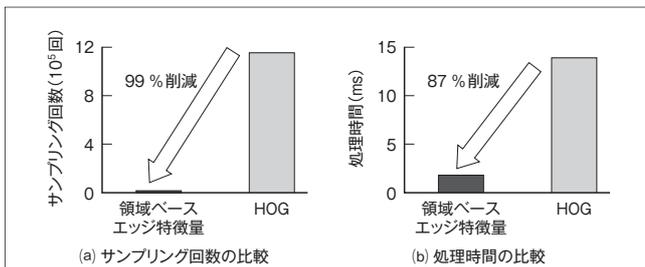


図5. 予備実験の結果 — 開発した領域ベースエッジ特徴量は、従来のHOGに比べて、サンプリング回数が少なく処理時間も短い。
Results of preliminary experiments on region-based edge feature quantity calculation

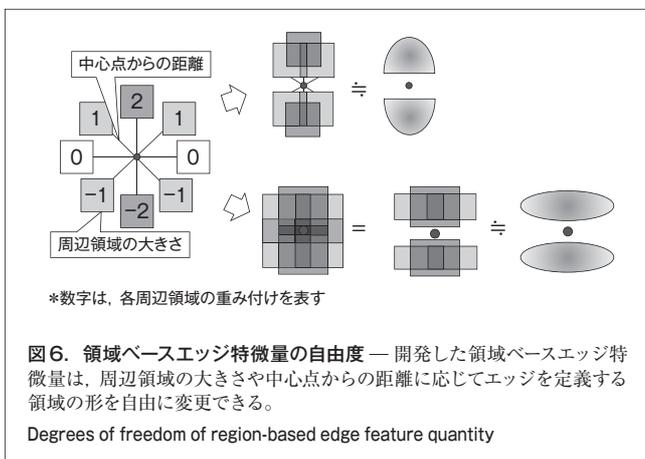


図6. 領域ベースエッジ特徴量の自由度 — 開発した領域ベースエッジ特徴量は、周辺領域の大きさや中心点からの距離に応じてエッジを定義する領域の形を自由に変更できる。
Degrees of freedom of region-based edge feature quantity

高速な特徴量の算出が可能であることを確認した。

開発した領域ベースエッジ特徴量は、周辺領域の大きさや中心点からの距離に応じて画素情報の取得範囲や範囲内の画素の重み付けを変えることができる。図6に示すように、縦方向のエッジは周辺領域内の輝度値合計の重み付き和によって算出できる。例えば、中心点からの距離を近づけ周辺領域を大きくすることで、取得範囲を半円形状やだ円形状に近似するとともに、取得範囲の中心付近の画素の重み付けを大きくできる。このように、領域ベース特徴量は、単純に画像縮小して隣接画素のエッジを算出することだけでは得られない特徴量を設定できるため、高い認識性能を得ることができる。

3 実験結果

かざすユーザーインタフェース向け画像認識技術の効果を実験結果で示す。登録する画像数を増加させた場合と、領域ベースエッジ特徴量を算出する際の周辺領域の大きさと中心からの距離を変えた場合について認識性能を評価した。

まず、登録する画像の枚数を増加させた場合の実験結果について述べる。ここでは700種類の雑誌について、正面から撮影した条件である面内回転角0°の画像をシステムに登録し、面内回転角0°、10°、及び20°の斜めから撮影した条件の画像を入力データとして用いた。認識性能として、処理時間、メモリ使用量、及び認識率について、従来手法のORB及びHOGと比較した。認識率の実験結果を表1に、処理時間と

表1. 認識率の比較

Comparison of recognition rates of conventional edge and region-based edge feature quantity calculations

手法	領域ベースエッジ特徴量	ORB	HOG
認識率 (%)	99	99	89

*1: 認識率 = 100 - 等価エラー率 (%)
*2: 等価エラー率は、他の対象物とまちがえてしまう割合 (False Acceptance Rate) とその対象物なのに違うと認識する割合 (False Reject Rate) が等しいときの割合

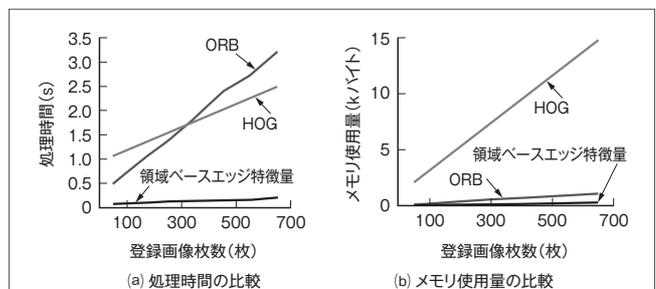


図7. 処理時間とメモリ使用量の比較 — 開発した領域ベースエッジ特徴量は、従来のHOGとORBに比べ、処理時間とメモリ使用量ともに優れている。
Comparison of response times and memory access costs of conventional edge and region-based edge feature quantity calculations

メモリ使用量の実験結果を図7に示す。処理時間は Intel® Core™ (注2) i5プロセッサ (動作周波数: 2.67 GHz) のパソコンで測定した。HOGは回転方向の変化に弱いため、領域ベースエッジ特微量やORBに比べて認識率が低いことがわかる。ORBは局所領域を使う特微量の中では高速な手法として知られているが、登録物体数が増加するとともに処理時間が大きく増えていることがわかる。これは、登録画像データベースとのマッチングの際に、ORBを算出する局所領域どうしの比較回数が増大するからである。一方、領域ベースエッジ特微量は、登録画像枚数が増加した場合でも、少ないメモリ使用量で良好な応答速度と高い認識率を実現できることがわかる。

次に、領域ベースエッジ特微量を算出する際の周辺領域の大きさと中心からの距離を変えた場合について述べる。ここでは、単純に画像縮小して隣接画素のエッジを算出する場合と比較した。図8に示す実験結果から、周辺領域の大きさや中心からの距離に自由度を持つ領域ベースエッジ特微量のほうが、認識率はより高くなることがわかった。

最後に、開発した領域ベースエッジ特微量をタブレットで動作させた場合の実験結果を示す。使用したタブレットはCPUにARM Cortex™ (注3)-A9 (動作周波数: 1.3 GHz) を搭載しており、1枚の画像を入力したときの処理時間は340 msであった。更に、この領域ベースエッジ特微量を用いた画像認識技術をかざすユーザーインターフェースに組み込んだ二つの例を図9に示す。図9(a)は、気になった雑誌にかざすことで関連する雑誌のバックナンバーを提示するインターフェースである。また図9(b)は、子どもが描いたスケッチにかざすことで、そのときの思い出の写真を提示するユーザーインターフェースである。

これらの結果より、開発した領域ベースエッジ特微量による画像認識技術は、携帯端末を用いて良好な応答速度と高い認識性能で動作させることができることから、ユーザビリティに優れたユーザーインターフェースの実現に貢献できると考えられる。

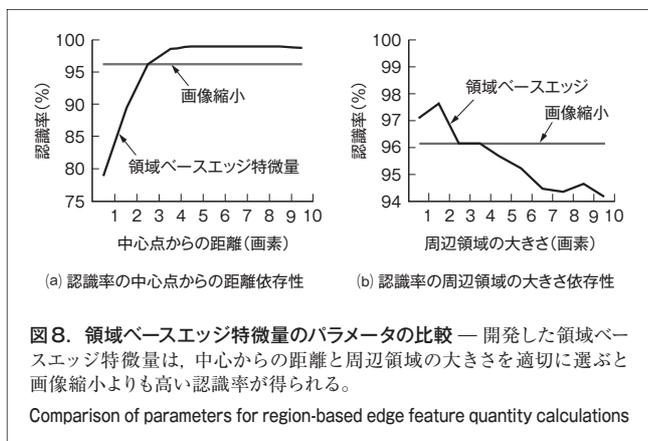


図8. 領域ベースエッジ特微量のパラメータの比較 — 開発した領域ベースエッジ特微量は、中心からの距離と周辺領域の大きさを適切に選ぶと画像縮小よりも高い認識率が得られる。

Comparison of parameters for region-based edge feature quantity calculations

(注2) Intel, Intel Coreは、米国及びその他の国におけるIntel Corporationの商標。

(注3) ARM, Cortexは、英国ARM Limitedの商標。



図9. かざすユーザーインターフェースの例 — 気になった雑誌や子どもが描いたスケッチにタブレットをかざすだけで、それぞれの関連情報が提示される。

Examples of attention-focusing UI installed on tablet

4 あとがき

身の回りにある実世界の物からクラウドシステム上の豊富なコンテンツへ、シームレスにつながる新しいユーザーインターフェースを目指して、そのコアとなる画像認識技術について述べた。開発した技術によって、スマートフォン上で約0.5秒の処理時間で700枚の登録画像に対して99%の認識率で対象物を見分けることが可能になった。

また、この技術は、身近なところで“きっかけ”を与えることができる点がポイントであり、そのきっかけに対する驚き、おもしろさや手軽さなどをユーザーに提供する新しいサービスを、展開していきたいと考えている。

文献

- (1) Rublee, E. et al. "ORB: An efficient alternative to SIFT or SURF". 2011 IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, 2011-11, IEEE, 2011, p.2564 - 2571.
- (2) Dalal, N. et al. "Histograms of oriented gradients for human detection". IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005 (CVPR 2005), San Diego, CA, USA, 2005-06, IEEE, 2005, p.886 - 893.



西山 正志 NISHIYAMA Masashi, Ph.D.

研究開発センター インタラクティブメディアラボラトリー研究主務、博士(学際情報学)。拡張現実感及び画像認識の研究・開発に従事。電子情報通信学会会員。

Interactive Media Lab.



大平 英貴 OHIRA Hidetaka

研究開発センター インタラクティブメディアラボラトリー。拡張現実感及び画像認識の研究・開発に従事。

Interactive Media Lab.



関根 真弘 SEKINE Masahiro

研究開発センター インタラクティブメディアラボラトリー研究主務。拡張現実感及びコンピュータグラフィックスの研究・開発に従事。映像情報メディア学会会員。

Interactive Media Lab.